LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-646114

X.XX Catalyst Advances Big Data, HPC Simulation, and Technology Partnerships

M. L. Leininger

November 12, 2013

Computation Annual Report 2013

Disclaimer

X.XX Catalyst Advances Big Data, HPC Simulation, and Technology Partnerships

Executive Summary [20% of article]

Lawrence Livermore National Laboratory (LLNL) has multiple mission areas, including stockpile stewardship, bioinformatics, nonproliferation, counterterrorism, and many broader national security programs. Each of these mission areas is proving to be in need of data science and analytics capabilities in addition to, or in conjunction with LLNL's high performance computing (HPC) simulation capabilities. LLNL in collaboration with two U.S. industry partners, Intel and Cray, have deployed a one-of-a-kind system, called Catalyst, to explore the system software and application algorithms for addressing future data analytics and modeling issues. Through LLNL's HPC Innovation Center other U.S. companies can also use Catalyst to investigate and develop solutions for industrial applications.

Progress in 2013 [80% of article]

For over a decade, Livermore Computing has successfully pushed the state of the art in high-performance computing (HPC) cluster technologies by transforming Linux clusters from small departmental systems into large-scale work horse platforms that are capable of running our most challenging scientific and engineering simulations. Today these simulations can generate terabytes to petabytes of data leading to a "big data" analysis problem. Big data can also arise in data collection from experiments, sensor or social networks, biological system studies, and many other sources. To explore the future needs of HPC and big data, Lawrence Livermore in partnership with Intel and Cray have created a unique system called Catalyst.

The 150 teraflop/s Catalyst cluster has 324 nodes, 7,776 CPU cores, and employs the latest generation Intel 12-core Xeon E5-2695v2 processors. Catalyst runs the NNSA-funded Tri-lab Open Source Software (TOSS) that provides a common user environment across NNSA Tri-lab clusters. Catalyst features include 128 gigabytes (GB) of dynamic random access memory (DRAM) per node, 800 GB of non-volatile memory (NVRAM) per compute node, 3.2 terabytes (TB) of NVRAM per Lustre router node, and improved cluster networking with dual rail Quad Data Rate (QDR-80) Intel TrueScale.

The Catalyst architecture is expected to provide insights into the kind of technologies the Advanced Simulation and Computing (ASC) program will require over the next 5-10 years to meet high performance simulation and big data computing mission needs. The increased storage capacity of the system (in both volatile and nonvolatile memory) represents the major departure from classic simulation-based computing architectures common at DOE laboratories and opens new opportunities for exploring the potential of combining floating point focused capability with data analysis in one environment. Consequently, the insights provided by Catalyst could become a basis for future commodity technology procurements.

The Catalyst collaboration involves two US companies partnering with LLNL to explore the use of a one-of-a-kind system. The partners will take advantage of Catalyst to explore different approaches and algorithms for addressing a wide variety of data analytics and modeling issues both inside and outside of the DOE mission space. In addition, by partnering with LLNL's HPC Innovation Center companies can also investigate the utility of machines like Catalyst for developing solutions to industrial applications. These essential collaborations accelerate innovation by bringing together hardware and software engineers, computational scientists, applied mathematicians and computer scientists to develop next-generation solutions for pressing problems in such diverse areas as bioinformatics, energy distribution and national security.

The HPC Innovation Center at LLNL will offer access to Catalyst and the expected big data innovations it enables as new options for its ongoing collaborations with American companies and research institutions. Catalyst will extend the range of possibilities for the processing, analysis and management of the ever larger and more complex datasets that many areas of business and science now confront.

Word-Count Reminder:
1-page article: 1 figure = 550 words
2-page article: 2 figures = 1000 words; 1 figure = (2-page article should have 2 figures)
The text below this box is not included in the word count.

DC Review: DO NOT SUBMIT ARTICLE WITHOUT COMPLETING THIS SECTION

Name of Derivative Classifier or DUSA: Date Reviewed:

Topics/Guides:

Contact information

Name, email@llnl.gov
External Web site (if applicable), website.llnl.gov

Captions

Figure a. Catalyst is a unique high performance computing (HPC) cluster that will serve research scientists and provide a proving ground for new HPC and Big Data technologies and architectures.

Prepared by LLNL under Contract DE-AC52-07NA27344.